

目 次

第 1 章 導 入	1
1.1 自然言語処理の困難さ	1
1.2 ニューラルネットワークと深層学習	3
1.3 自然言語処理における深層学習	3
1.3.1 成功事例	5
1.4 本書で扱う内容とその構成	7
1.5 本書で扱っていない内容	10
1.6 用語に関する注意	11
1.7 数学的記法について	11
第 1 編 教師あり分類とフィードフォワードニューラルネットワーク	13
第 2 章 機械学習の基礎と線形モデル	15
2.1 教師あり学習とパラメータを持つ関数	15
2.2 訓練セット, テストセット, 検証セット	17
2.3 線形モデル	19
2.3.1 ニクラス分類	19
2.3.2 対数線形ニクラス分類	23
2.3.3 多クラス分類	24
2.4 表現	25
2.5 ワンホットベクトル表現と密ベクトル表現	26
2.6 対数線形多クラス分類	28

目 次

2.7	最適化としての訓練	28
2.7.1	損失関数	29
2.7.2	正則化	33
2.8	勾配に基づく最適化	34
2.8.1	確率的勾配降下法 (SGD)	36
2.8.2	事例詳説	37
2.8.3	SGD を超えて	39
第 3 章	線形モデルから多層パーセプトロンへ	41
3.1	線形モデルの限界：排他的論理和問題	41
3.2	非線形入力変換	42
3.3	カーネル法	43
3.4	訓練可能な写像関数	43
第 4 章	フィードフォワードニューラルネットワーク	45
4.1	脳にヒントを得た比喻	45
4.2	数学的記法による記述	47
4.3	表現力	49
4.4	一般的な非線形要素	50
4.5	損失関数	52
4.6	正則化とドロップアウト	52
4.7	類似度と距離の層	53
4.8	埋め込み層	54
第 5 章	ニューラルネットワークの訓練	57
5.1	計算グラフによる抽象化	57
5.1.1	前向き計算	59
5.1.2	後向き計算（導関数と逆伝播）	60
5.1.3	ソフトウェア	61
5.1.4	実装方法	64
5.1.5	ネットワークの合成	65
5.2	実践豆知識	65
5.2.1	最適化アルゴリズムの選択	66
5.2.2	初期化	66
5.2.3	リスタートとアンサンブル	67
5.2.4	勾配消失と勾配爆発	67
5.2.5	飽和ニューロンと死亡ニューロン	68
5.2.6	シャッフリング	68

5.2.7	学習率	68
5.2.8	ミニバッチ	69
第 2 編	自然言語データの扱い	71
第 6 章	テキストデータのための素性	73
6.1	自然言語処理における分類問題のタイプ分け	74
6.2	自然言語処理問題のための素性	76
6.2.1	直接観察できる属性	76
6.2.2	推測される言語学的特徴	80
6.2.3	核となる素性と組み合わせ素性	85
6.2.4	n -グラム素性	86
6.2.5	分布論的素性	87
第 7 章	事例研究：自然言語処理における素性	89
7.1	文書分類：言語同定	89
7.2	文書分類：トピック分類	90
7.3	文書分類：著者特定	91
7.4	文脈に埋め込まれた単語：品詞タグ付け	91
7.5	文脈に埋め込まれた単語：固有表現認識	93
7.6	文脈に埋め込まれた単語と言語学的素性：前置詞意味曖昧性解消 ...	95
7.7	文脈に埋め込まれた単語の間の関係： アークを単位としたパーズング	98
第 8 章	テキストの素性から入力への変換	101
8.1	カテゴリ素性の符号化	101
8.1.1	ワンホット符号化	102
8.1.2	密な符号化（素性埋め込み）	102
8.1.3	密ベクトル表現とワンホット表現	104
8.2	密ベクトルの組み合わせ	105
8.2.1	窓に基づく素性	105
8.2.2	可変数の素性：連続的単語バッグ (CBOW)	106
8.3	ワンホットベクトルと密ベクトルの関係	107
8.4	諸事項	108
8.4.1	距離素性と位置素性	108
8.4.2	パディング, 未知語, 単語ドロップアウト	109
8.4.3	素性の組み合わせ	111

目 次

8.4.4	ベクトル共有	112
8.4.5	次元数	112
8.4.6	埋め込みの語彙	113
8.4.7	ネットワークの出力	113
8.5	事例紹介：品詞タグ付け	113
8.6	事例紹介：アークを単位としたパーズング	115
第 9 章	言語モデリング	119
9.1	言語モデリングタスク	119
9.2	言語モデルの評価：パープレキシティ	121
9.3	言語モデリングの古典的手法	121
9.3.1	補足的文献	123
9.3.2	古典的言語モデルの限界	123
9.4	ニューラル言語モデル	124
9.5	生成における言語モデルの利用	128
9.6	副産物：単語表現	129
第 10 章	事前学習された単語表現	131
10.1	無作為初期化	131
10.2	タスクに固有の教師あり事前学習	132
10.3	教師なし事前学習	132
10.3.1	事前学習された埋め込みの利用	133
10.4	単語埋め込みアルゴリズム	134
10.4.1	分布仮説と単語表現	135
10.4.2	ニューラル言語モデルから分散表現へ	140
10.4.3	二つの世界の関係	144
10.4.4	その他のアルゴリズム	145
10.5	文脈の選択	146
10.5.1	窓に基づく方法	146
10.5.2	文、パラグラフ、文書	147
10.5.3	統語的な窓	147
10.5.4	多言語	149
10.5.5	文字に基づく表現とサブワード表現	149
10.6	連語と屈折の扱い	151
10.7	分布に基づく手法の限界	152
第 11 章	単語埋め込みの利用	155
11.1	単語ベクトルの獲得	155

11.2	単語の類似度	156
11.3	単語のクラスタリング	156
11.4	類義語の発見	157
11.4.1	語群との類似度	157
11.5	仲間外れ探し	158
11.6	短い文書の類似度	158
11.7	単語の類推	159
11.8	レトロフィッティングと射影	160
11.9	実践における落とし穴	161
第 12 章	事例研究：文の意味推論のためのフィードフォワードアーキテクチャ ..	163
12.1	自然言語推論と SNLI データセット	164
12.2	テキストの類似性を判定するネットワーク	165
第 3 編	特別なアーキテクチャ	169
第 13 章	n-グラム検出器：畳み込みニューラルネットワーク	173
13.1	基本的な畳み込みとプーリング	175
13.1.1	テキストについての 1 次元畳み込み	176
13.1.2	ベクトルプーリング	178
13.1.3	変種	181
13.2	代替法：素性ハッシュ	181
13.3	階層的畳み込み	182
第 14 章	再帰的ニューラルネットワーク：系列とスタックのモデリング	187
14.1	RNN 抽象化	188
14.2	RNN の訓練	191
14.3	一般的な RNN の利用方法	191
14.3.1	受容器（アクセプタ）	191
14.3.2	符号化器（エンコーダ）	192
14.3.3	変換器（トランスデューサ）	193
14.4	双方向 RNN (biRNN)	194
14.5	多層 RNN（積み上げ RNN）	196
14.6	RNN によるスタックの表現	197
14.7	文献を読む際の注意	199
第 15 章	RNN の具体的な構成	201
15.1	RNN としての CBOW	201

目次	
15.2	単純 RNN 202
15.3	ゲート付きアーキテクチャ 202
15.3.1	LSTM 204
15.3.2	GRU 206
15.4	その他の変種 207
15.5	RNN におけるドロップアウト 208
第 16 章	RNN を用いたモデリング 211
16.1	受理器 (アクセプタ) 211
16.1.1	感情分類 212
16.1.2	主語動詞一致についての文法性判定 214
16.2	素性抽出器としての RNN 216
16.2.1	品詞タグ付け 216
16.2.2	RNN-CNN 文書分類 218
16.2.3	アークを単位とした依存構造パーズング 219
第 17 章	条件付き生成 221
17.1	RNN 生成器 221
17.1.1	生成器の訓練 222
17.2	条件付き生成 (符号化器復号化器) 223
17.2.1	系列-系列モデル 225
17.2.2	応用 225
17.2.3	その他の条件付けの文脈 229
17.3	文の類似度の教師なし学習 230
17.4	アテンションあり条件付き生成 231
17.4.1	計算の複雑さ 235
17.4.2	解釈のしやすさ 235
17.5	自然言語処理におけるアテンションに基づくモデル 236
17.5.1	機械翻訳 236
17.5.2	形態論的屈折 238
17.5.3	統語パーズング 238
第 4 編	追加的な話題 239
第 18 章	RecNN による木構造のモデリング 241
18.1	形式的定義 242
18.2	拡張と変種 245

18.3	木構造ニューラルネットワークの訓練	245
18.4	シンプルな代替案：木を線条化する	246
18.5	今後の見通し	247
第 19 章	構造を持つ出力の予測	249
19.1	探索に基づく構造予測	249
19.1.1	線形モデルによる構造予測	250
19.1.2	非線形な構造予測	250
19.1.3	CRF による確率的目的関数	252
19.1.4	近似探索	253
19.1.5	再ランキング	254
19.1.6	さらなる参照	255
19.2	貪欲法による構造予測	255
19.3	構造を持つ出力の予測としての条件付き生成	256
19.4	例	258
19.4.1	探索に基づく構造予測：一次の依存構造パーズング	258
19.4.2	固有表現認識のためのニューラル CRF	259
19.4.3	ビーム探索による近似的 NER-CRF	263
第 20 章	モデルのカスケード接続, マルチタスク学習, 半教師あり学習	265
20.1	モデルのカスケード接続	266
20.2	マルチタスク学習	270
20.2.1	マルチタスクの設定における訓練	272
20.2.2	選択的共有化	272
20.2.3	マルチタスク学習としての単語埋め込みの事前学習	273
20.2.4	条件付き生成におけるマルチタスク学習	274
20.2.5	正則化としてのマルチタスク学習	274
20.2.6	いくつかの警告	274
20.3	半教師あり学習	275
20.4	様々な例	276
20.4.1	視線予測と文圧縮	276
20.4.2	アークへのラベル付けと統語パーズング	278
20.4.3	前置詞の意味曖昧性解消と翻訳	279
20.4.4	条件付き生成：多言語機械翻訳, パーズング, 画像キャプション生成	280
20.5	今後の見通し	282
第 21 章	結 論	283

目 次

21.1	これまでの進展	283
21.2	これからの課題	283
文献一覧		285
訳者あとがき		311
略語一覧		314
索 引		315